SCI 数据通信技术的研究和实现

余强力,陆 达,王明芬
(厦门大学 计算机系,福建 厦门 361005)

摘 要:随着计算机应用领域的拓展与加深,对计算机性能的要求也越来越高,进而对系统中的信息交互提出了更高的要求。传统的总线技术由于存在固有的限制,已经成为阻碍系统性能提高的瓶颈。SCI 的出现克服了总线技术的不足,大大改善了信息交互的效率和带宽,满足了系统中和系统间的互连要求,成为先进互连技术的代表。详细分析了在 Windows XP 环境下,开发 SCI 数据通信程序的原理和方法。

关键词:SCI;数据通信;接口
中图分类号:TP311.52 **文献标识码**:A **文章编号**:1673-629X(2007)11-0190-03

Research and Implementation on SCI Data Transmission

YU Qiang-li, LU Da, WANG Ming-fen
(Computer Science Department, Xiamen University, Xiamen 361005, China)

Abstract: With the development of computer technology, the requirement for the performance of computer and transmission become more and more high. Because of the limitation of traditional bus technology, it becomes the choke point of advance the performance of system. SCI overcomes the limitation, enhances the efficiency and bandwidth of transmission. It meets the demand of current transmission, becomes the mainstream of transmission technology. The architecture of SCI device driver in Windows XP environment is described in this paper and data transmission methods on SCI network are also presented with examples.

Key words: SCI; data transmission; interface

0 引 言

Scalable Coherent Interface (SCI,可扩展的一致性接口)是一种高性能系统互连技术,可以用来构建不同拓扑结构的多处理器系统。SCI 采用单向的点到点传输机制从而避免了总线结构的局限性,大大改善了信息交互的效率和带宽,满足了系统中和系统间的互连要求。其低延迟、高带宽的特点完全可胜任各种关键应用领域。文中给出 SCI 数据通信软件的设计。

1 SCI 概述

SCI[1~3]是一种系统互连技术。它定义了 3 个子系统:物理层接口、基于数据包的逻辑通信协议以及分布式缓存一致性协议。其中,物理层接口定义了诸如板级大小、介质连接、网络时钟等特性。而通信协议是基于单向点到点的连接。另外,为了优化系统访问的延迟特性,SCI 支持存储器分级策略,支持高速缓存机制,并采用分布式链式目录的方法来有效地支持缓存机制。

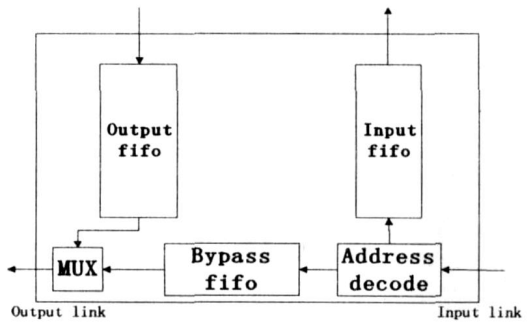
SCI 可支持 DSM(分布式共享存储器)体系结构。具有 64 位地址,其中前 16 位用于节点地址分配,后 48 位用于每个节点内部的地址分配。也就是说 SCI 最多可支持 65536 个节点的互连。采用 DSM 体系结构可以提供良好的扩展性,系统规模可随着资源的增加非常方便地扩展。带宽为 18bit 的 18-DE-500 并行链路,采用差分信号传输,每条信号线可提供 500Mbit/s 的带宽,最高可支持 1Gbyte/s 的并行传输。SCI 支持灵活的拓扑结构,基本的 SCI 拓扑是环。与传统的环形拓扑不同的是它可支持消息的并发传输。

如图 1 所示,基本 SCI 接口包括:一个旁路 FIFO,一个接收单元和一个发送单元。其中,为了实现并行发送和接收消息,接收单元和发送单元各自有一个队列,从而避免数据包交互时的死锁。因为只有当旁路 FIFO 为空时节点才能够传输消息,旁路 FIFO 的大小取决于节点能够产生的最大数据包。在包的发送期间,可以利用空闲 symbol 排空旁路 FIFO。另外,Input 和 Output FIFO 能够匹配节点处理速度与链路传输速

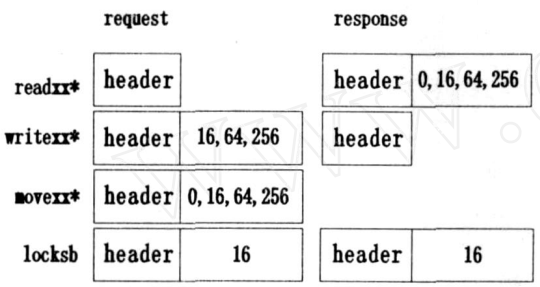
收稿日期:2007-01-30
作者简介:余强力(1983-),男,福建福清人,硕士研究生,研究方向为多计算机系统通信;陆 达,教授,硕士生导师,研究方向为计算机控制。

度之间的差异。

SCI 链路上传输信号的最小单位是 symbol。一个 symbol 包括 16 位的数据位,包分割符以及时钟信息。连续的 symbol 构成一个包。另外,在传输间隔也可以通过发送空闲符号(idle symbol)维护链路和接收方的同步。如图 1 所示,SCI 提供的事务类型包括 read, write,lock 和 move。其中前三种事务中的请求过程会有相应的响应确认。而 move 不需要响应确认。因此在可靠性要求不高的情况下能提高传输的效率。



(a)　SCI 物理接口



(b)　事务类型

图 1　SCI 物理接口和事务类型

2　SCI 在 Windows XP 下的驱动模型

SCI 支持各种主流的操作系统,文中的编程是在 Windows XP 下实现的。目前 SCI 接口卡的主要厂家是挪威的 Dolphin 公司,文中采用的是该公司的 Intel 平台上 PCI 总线的 D331 SCI 接口卡[4]。用它构造的通信系统的开发环境如图 2 所示。

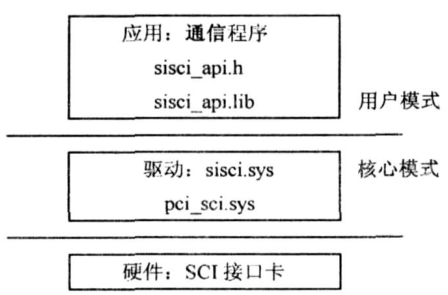


图 2　Windows XP 下 SCI 驱动模型

pcisci.sys 完成 PCI 总线设备的访问功能,并且将 PCI 总线事务映射成为 SCI 网络事务,提供透明的 SCI 设备访问机制。sisci.sys 提供了 SCI 网络的高层驱动,它屏蔽了 SCI 协议的细节,为应用系统设计者提供了数据通信接口。sisci_api.h 和 sisci_api.lib 为 Win32 应用程序提供了 SISCI API 函数的调用窗口。

3　SCI 在 Windows XP 下通信程序的实现

首先需要了解两个重要概念:

(1)虚拟设备(virtual device),可以把它看成是在 SCI 驱动模型下进行通信的管道,所以在运行其它 SISCI API 函数前,必须创建虚拟设备。SCIOpen(&sd,NO_FLAGS,&error);//创建虚拟设备。

(2)内存段(memory segment),能够异地访问其它节点上的本地内存是 SCI 的基本特点。

下面要介绍的 PIO 方式和 DMA 方式都需要处理好管理本地内存段和连接远端内存段的工作。图 3 是共享内存段通信的结构图。两种通信方式都是基于以下结构实现的[5]。

SCI 提供了中断方式来实现 SCI 节点间的事件通知。中断也分为本地中断和远端中断,Server 节点建立本地中断,并在中断上等待。Client 节点连接 Server 节点上的中断,然后触发中断,最后在不用的时候断开中断连接。

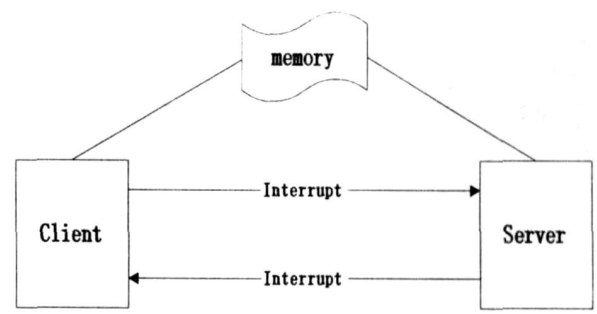


图 3　共享内存段通信结构图

下面以两个节点通信为例介绍两种通信方式:一个发送数据的节点叫 Client,另一个接收数据的节点叫 Server,它们之间通信过程是以中断方式来协调的。

3.1　PIO 方式

远端内存段映射到本地进程地址空间之后,本地进程就可以使用内存复制命令来完成数据传输,这种方式叫 Program IO(PIO),因为这个复制过程是由 CPU 来完成。它是针对小规模数据传输的一种通信方式。Server 节点声明 sci_local_segment_t 型变量 localseg_s,然后用这个变量使本地内存映射到 SCI 全局地址空间,使之为其它节点可见:

```
sci_local_segment_t localseg_s;
sci_map_t local_map;
```

```
int * local_address;
SCICreateSegment( …, &localseg_s, …) ;// 创建一个本地内存段
Local_address = (volatile int *) SCIMapLocalSegment (localseg_s, &local_mao, …) // 映射到本地进程地址空间
```

把本地内存段映射到 SCI 全局地址空间,并使它为其它节点可见:

```
SCIPrepareSegment(localseg_s, …) ;
SCISetSegmentAvailable(localseg_s) ;
```

Client 节点声明一个 sci_remote_segment_t 型变量 remoteseg_c,用它来连接 Server 节点共享的内存段,这样就可以用内存复制命令来完成数据传输:

```
sci_remote_segment_t remoteseg_c;
sci_map_t remote_map;
int * remote_address;
SCIConnectSegment( …, &remoteseg_c, …) ;// 连接 Server 节点共享的内存段
remote_address = (volatile int *) SCIMapRemoteSegment (remoteseg_c, &remote_map, …) ;// 映射 Server 节点的共享内存段到本地进程地址空间
```

同样,在程序结束时不忘断开连接,释放资源:

```
SCIDisconnectSegment(remoteseg_c, …) ;
SCIRemoveSegment( …) ;
```

3.2 DMA 方式

在 DMA 方式中,数据从一个节点复制到另一个节点不是由 CPU 来执行的,而是由 SCI 接口卡上的 DMA 引擎来完成的。所以 DMA 方式允许在数据传输时,CPU 做其它的事情。适合大量的数据传输。

DMA 方式中的 Server 节点的设计跟共享内存方式差不多:创建一个内存段,把它映射到 SCI 全局地址空间,使其它节点可见。下面主要介绍 Client 节点的设计:

```
SCICreateSegment() ;// 创建本地内存段
SCIMapLocalSegment() ;// 映射本地内存段到本地进程地址空间
SCICreateDMAQueue() ;// 创建 DMA 队列
SCIConnectSegment() ;// 连接 Server 节点的共享内存段
/ *把要发送的数据,写入本地内存段 */
SCIEnqueueDMATransfer() ;// 将要发送的数据送到 DMA 队列
SCIPostDMAQueue() ;// 把队列中的数据投递给 DMA 引擎
SCIWaitForDMAQueue() ;// 等待 DMA 传输完成
SCIDMAQueueState() ;// 查询 DMA 队列的状态,以确定正常完成
```

最后,通信结束后,删除队列,释放资源:SCIRemoveDMAQueue() ;。

4 结 论

从测试的数据上可以看出,PIO 在小数据量(几十 kB)的传输上性能明显比 DMA 好,这是因为 DMA 方式需要时间来创建 DMA 引擎,而且还要管理 DMA 队列,所以会增加一点延迟。不过随着数据量的增加,DMA 的优势就显示出来了。

现在 SCI 技术已经在许多高性能处理器互连方面得到广泛的应用,其他一些要求高性能的 IO 领域也在展开应用研究。文中给出了在 Windows XP 下,SCI 的数据通信程序的详细实现方法。目前 SCI 只提供同步的通信方式,所以在实际应用中,可以用 Windows XP 的多线程编程技术来实现异步通信。

参考文献:

[1] IEEE Std 159621992. Scalable Coherent Interface(SCI) [S]. USA: The Institute of Electrical and Electronics Engineers, Inc,1993.

[2] 蔡敏芳,陆 达. SCI 协议在高性能集群计算中的可行性分析[J]. 计算机科学与实践,2005:3(8) :4 - 8.

[3] 周 强,罗志强. SCI 协议标准综述[J]. 航空电子技术,2001(6) :9 - 18.

[4] Dolphin Interconnect Solutions AS. Installation Guide for the D33x2family PCI2SCI Adapter Card PCI642bit/ 66MHz SCI AdapterCard [R]. Norway: Dolphin Interconnect Solutions AS,2001.

[5] Dolphin Interconnect Solutions AS. SISCI API User Guide [R]. Norway:Dolphin Interconnect Solutions AS,2001.

(上接第 189 页)

业知识管理关键技术仍处于探索阶段,还有待今后作进一步研究和完善。

参考文献:

[1] 禅 影. 新版会议须知 ——读《会议管理 ——如何创造高效率会议》[J]. 软件工程师,2003(3) :59 - 60.

[2] 林 山,黄培伦. 基于企业知识理论的组织创新探索[J]. 中国科技论坛,2004(11) :37 - 40.

[3] Nanaka, Takeuchi H. The knowledge creating company: How Japanese companies create the dynamics of innovation [M]. New York:Oxford University Press,1995.

[4] 李牧原,佘素丽. 知识基础的企业创新模型[J]. 兰州学刊,2004(1) :89 - 90.

[5] Bhatt G D. Organizing in the Knowledge Development Cycle [J]. Journal of Knowledge Management,2000,4(1) :15 - 26.

[6] 徐瑞平,王 丽,陈菊红. 基于知识价值链的企业知识创新动态模式研究[J]. 科学管理研究,2005(8) :78 - 81.